

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Кафедра «Програмної інженерії та інформаційних  
технологій управління»

МЕТОДИЧНІ ВКАЗІВКИ

до лабораторної роботи № 1 за курсом

**«Інтелектуальний аналіз даних та видобування знань»**

для студентів спеціальностей : спеціальності 121- Інженерія програмного  
забезпечення, спеціалізації 121 – 01 - Інженерія програмного забезпечення

Затверджено

редакційно-видавничою радою університету,

протокол №      від      2017 року

Харків НТУ «ХПІ» 2017

Методичні вказівки до проведення лабораторних занять за курсом "Інтелектуальний аналіз даних та видобування знань" для студентів спеціальності 121- Інженерія програмного забезпечення, спеціалізації 121 – 01 Інженерія програмного забезпечення \ Уклали – Єршова С. І., Мельник К. В. – Харків: НТУ «ХПІ», 2017. – с.

Укладачі С. І. Єршова, К. В. Мельник

Рецензент О. Ю. Чередніченко

Кафедра програмної інженерії та інформаційних технологій управління

## ЗМІСТ

|   |    |
|---|----|
| Вступ.....  | 4  |
| Загальні положення.....                                       | 5  |
| Правила виконання і оформлення лабораторних робіт .....       | 6  |
| Лабораторна робота № 1 «Моделювання предметної області» ..... | 10 |
| Контрольні питання .....                                      | 18 |
| Література .....  | 19 |

## ВСТУП

Методичні вказівки містять завдання для виконання лабораторних робіт для студентів базового напрямку «Програмна інженерія» рівня магістрів.

Мета навчальної дисципліни є опанування студентами теоретичних основ інтелектуального аналізу даних та видобування знань, надання студентам систематизованих знань про основні моделі, методи та технології, які використовуються для інтелектуального аналізу даних та видобування знань; отримання студентами теоретичних знань та практичних навичок роботи з сучасними технологіями, отримання практичних навичок використання програмних засобів розробки онтологій у наукових дослідженнях; формування у студентів аналітичних здібностей, які б дозволили робити обґрунтований вибір методів, засобів та технологій при рішенні задач предметної області.

## ЗАГАЛЬНІ ПОЛОЖЕННЯ

Онтологія (від грец. онтос – суще, логос – навчання, поняття) – термін, що визначає вчення про буття, про сутність.

Термін ввів у філософську літературу німецький філософ Р. Гокленіус (1547–1628). До цього онтологія була частиною метафізики, наукою, що не пов’язана з логікою, з науками про природу. Її предмет становить вивчення абстрактних і загальних філософських категорій.

Сьогодні розуміння терміну «онтологія» різне, залежно від контексту і цілей його використання. У роботі [1] виділяють такі аспекти інтерпретації цього терміну:

- 1) онтологія як філософська дисципліна;
- 2) онтологія як неформальна концептуальна система;
- 3) онтологія як формальний погляд на семантику;
- 4) онтологія як специфікація “концептуалізації”;
- 5) онтологія як уявлення концептуальної системи через логічну теорію, що характеризується спеціальними формальними властивостями, її призначенням;
- 6) онтологія як словник;
- 7) онтологія як метарівнева специфікація логічної теорії.

Четверта інтерпретація, запропонована Томасом Грубером. Визначення найчастіше використовується сьогодні в роботах з проектування і дослідження онтологій.

## ПРАВИЛА ВИКОНАННЯ І ОФОРМЛЕННЯ ЛАБОРАТОРНИХ РОБІТ

Звіт до лабораторної роботи виконується на одній стороні аркуша білого паперу формату А-4 (210x297 мм) шрифтом Times New Roman розміру 14 через 1,5 міжрядний інтервал. На кожному аркуші повинні бути поля таких розмірів: ліворуч, зверху та знизу - не менше за 20 мм, праворуч - не менше ніж 10 мм. Шрифт друку повинен бути чітким, чорного кольору. Щільність тексту має бути однаковою.

Друкарські помилки, описки чи графічні нечіткості, виявлені під час оформлення роботи, можна виправляти охайним підчищенням чи за допомогою коректора і нанесенням на тому ж місці або між рядками виправленого тексту (фрагмента рисунка) тим самим кольором, яким написаний текст. Заголовки структурних частин роботи «ЗМІСТ», «ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ» (якщо такий є), «ВСТУП», «ХІД ВИКОНАННЯ РОБОТИ», «ВИСНОВКИ», «СПИСОК ДЖЕРЕЛ ІНФОРМАЦІЇ», «ДОДАТКИ» (якщо такі є) друкують великими літерами. Крапка наприкінці заголовка не ставиться.

Текст розділів може складатись з підрозділів. Заголовки підрозділів друкуються маленькими літерами (крім першої великої) з абзацного відступу. Крапка наприкінці заголовка не ставиться.

Якщо підрозділи містять пункти, то заголовки цих пунктів друкують маленькими літерами (крім першої великої) з абзацного відступу. Крапка наприкінці заголовка не ставиться..

Якщо заголовок складається із двох чи більше речень, то їх розділяють крапкою.

Нумерація сторінок, розділів, підрозділів, пунктів, підпунктів, рисунків, таблиць, формул подається арабськими цифрами без знака №.

Нумерація має бути наскрізною, причому першою сторінкою є титульний аркуш, але на ньому номер сторінки не ставиться. Рисунки і Таблиці включаються у загальну нумерацію. Сторінки з додатками і списком літератури

входять у наскрізну нумерацію. Номер сторінки проставляється у правому верхньому кутку сторінки без будь-яких знаків (крапки наприкінці, виділення рисками, дужками тощо).

Підрозділи нумеруються у межах розділу. Номер підрозділу складається із номера розділу та порядкового номера, розділених крапкою. У кінці також ставиться крапка, наприклад «2.1» (перший підрозділ (параграф) другого розділу). Пункти нумеруються арабськими цифрами у межах кожного підрозділу. Номер пункту має складатися із номера розділу, підрозділу і пункту (свого порядкового номера), розділених крапками. У кінці номера пункту ставиться крапка, наприклад, «2.1.3» (третій пункт першого підрозділу другого розділу). Потім у тому ж порядку йде заголовок пункту. Номери підрозділів і пунктів ставляться на їх початку, номер пункту на початку першого рядка абзацу, яким розпочинається відповідний пункт. Цифра номера пункту не повинна виступати за межі абзацу. Якщо пункти міститимуть підпункти, то їх нумерують у межах кожного пункту за такими ж правилами, як пункти.

Ілюстрації (фотографії, схеми, креслення, графіки тощо) називаються рисунками, які нумеруються послідовно у межах розділу або наскрізно арабськими цифрами: номер рисунку складається з двох цифр. Перша - номер розділу, друга - порядковий номер рисунку у розділі, розділених крапкою, наприклад «Рисунок 1.2» (другий рисунок першого розділу), а далі йде назва рисунка. При посиланні на рисунок перший раз необхідно вписувати його повний номер, наприклад, «(рис. 1.2)». Подальші посилання на рисунки виконуються разом із скороченим словом «дивись», наприклад «(див. рис. 1.2)». Рисунки мають розташовуватись одразу після посилання на них у тексті. Якщо на цій сторінці немає місця, то їх необхідно розташувати на наступній сторінці у зручній для ознайомлення формі тобто, щоб для вивчення цього рисунка сторінку можна було б повернути за годинниковою стрілкою. Номер рисунка, його назва і пояснювальні підписи розміщуються послідовно під рисунком.

Таблиці у роботі нумеруються послідовно (за винятком таблиць, поданих у додатках) в межах розділу. У правому верхньому куті пишеться слово «Таблиця» із зазначенням її номера. Номер таблиці складається із номера розділу і порядкового номера таблиці, розділених крапкою, наприклад «Таблиця 1.2» (друга таблиця першого розділу). При посиланні на таблицю вказують її повний номер, а слово «Таблиця» пишуть скорочено, наприклад «(табл. 1.2)». Якщо у роботі тільки одна таблиця, то її не нумерують і слово «Таблиця» не пишуть.

Під словом «Таблиця» розміщуються заголовки таблиці симетрично до форми таблиці. Слово «Таблиця» і заголовки починаються з великої букви. Назва не підкреслюється.

Таблиці потрібно розташовувати після першої згадки про неї у тексті так, щоб її зручно було розглядати без повороту або з поворотом за годинниковою стрілкою. Якщо на цій сторінці немає місця, то їх необхідно розташувати на наступній сторінці у зручній для ознайомлення формі, тобто, щоб для вивчення таблиці сторінку можна було б повернути за годинниковою стрілкою.

При переносі таблиці на наступну сторінку нумерацію колонок необхідно повторити, а над нею розмістити слова «Продовження табл.» із зазначенням її номера, наприклад «Продовження табл. 1.2».

Таблицю з великою кількістю рядків можна переносити на інший аркуш, при цьому назву розміщують тільки над її першою частиною. Таблицю з великою кількістю граф можна ділити на частини і розміщувати одну частину під іншою у межах одної сторінки. Якщо рядки або графи таблиці виходять за формат сторінки, то у першому випадку в кожній частині таблиці повторюють її головку.

Висота рядків повинна бути не меншою за 8 мм.

Формули в роботі (якщо їх більше від однієї) нумерують у межах розділу. Номер формули складається із номера розділу і порядкового номера формули у розділі, між якими ставиться крапка. Номер формули пишеться біля правого



краю аркуша на рівні відповідної формули (якщо формула велика, то на рівні нижнього рядка формули до якої він стосується) в круглих дужках, наприклад (2.2) (друга формула другого розділу). Пояснення позначень символів чи числових коефіцієнтів наводиться безпосередньо під формулою у тій послідовності, у якій вони подані у формулі. Це пояснення подається з нового рядка, починаючи зі слова "де", двокрапка після якого не ставиться. Значення кожного символу чи числового коефіцієнта подається з нового рядка. Розмірність одного і того самого параметра у межах роботи має бути однаковою. При посиланні у тексті на формулу, необхідно подати її повний номер у дужках, наприклад: «У формулі (1.2)».

Вище і нижче від кожної формули потрібно залишити не менше ніж один вільний рядок.

До списку джерел інформації входять усі використані джерела, які розташовуються за алфавітом або у послідовності посилань на них у тексті. Інформація про видання (монографії, підручники, довідники тощо) має містити: прізвище та ініціали автора, назву книги, місце видання, видавництво і рік видання, обсяг у сторінках.

Прізвище автора подається у називному відмінку. Якщо є два, три чи чотири автори, то їх прізвища з ініціалами подають у тій послідовності, у якій вони надруковані у книзі: перед прізвищем наступного автора ставлять кому. Якщо більше ніж чотири автори, вказують прізвища з ініціалами тільки перших трьох, а далі пишуть слова "та інші". Місце видання приводиться повністю у називному відмінку, можна скорочувати назви тільки : Київ (К. ).

Варіант задається за порядковим номером студента у журналі.

## ЛАБОРАТОРНА РОБОТА № 1

### «МОДЕЛЮВАННЯ ПРЕДМЕТНОЇ ОБЛАСТІ»

**Мета роботи:** Знайомство зі стандартом IDEF5 та його використання.

**Завдання:**

1. Провести аналіз предметної області згідно індивідуальному завданню, виділити окремі сутності та перелік значень. Перелік предметних областей наведено нижче. Номер області відповідає номеру у журналі академічної групи.
2. Розробити модель бази даних для обраної предметної області.
3. Розробити онтологію для обраної предметної області згідно стандарту IDEF5, обґрунтувати вибір типу схеми.

**Варіанти предметних областей до лабораторних робіт предметних областей:**

1. Діагностика стану автомобіля.
2. Управління пирососом, який сам їздить по квартирі і збирає пил.
3. Прогнозування пошиття верхнього одягу в залежності від погодних умов.
4. Оцінка кредитоспроможності клієнта.
5. Управління факультетом університету.
6. Управління тестуванням у ВУЗі.
7. Управління персоналом в організації в залежності від їх досягнень та поведінки в колективі.
8. Діагностика несправності комп'ютера.
9. Проектування озеленення міста в залежності від типу району: промисловий, житловий, змішаний.
10. Управління подачею води в багатоквартирному домі.
11. Діагностика несправності БФП (багатофункціонального приладу).

12. Управління поливом газонів в парку в місті.
13. Діагностика стану серцево-судинної системи людини.
14. Прогнозування поведінки арбітра при проведенні боксу.
15. Управління розробкою страхових програм.
16. Управління страховими договорами.
17. Прогнозування прибутку організації.
18. Управління якістю продукції в залежності від відгуків покупців.
19. Управління осушувачем повітря в спортзалі в залежності від умов в залі та погодних умов на вулиці.
20. Управління супутниковою антеною в залежності від погодних умов та відстані супутників від Землі, що транслюють певний канал.

### **Довідковий матеріал до лабораторної роботи:**

Стандарт онтологічного дослідження IDEF5 розроблений фірмою Knowledge Base Systems [2].



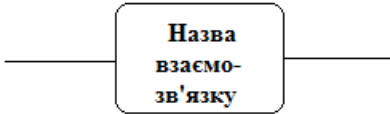

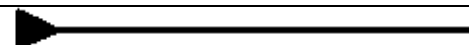
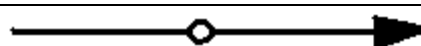
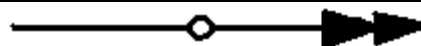


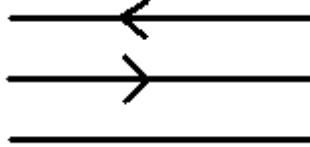

Процес розробки онтології в рамках стандарту IDEF5 складається з 5 етапів:

1. Вивчення і систематизація початкових умов. На цьому етапі визначається мета розробки онтології.
2. Збір і накоплення даних для побудови онтології.
3. Аналіз і групування даних, узгодження термінології.
4. Формування попередньої термінології.
5. Уточнення і затвердження онтології.

Для підтримки процесу побудови онтологій в IDEF5 існують спеціальні онтологічні мови: схематична мова (Schematic Language-SL) і мова розробки і удосконалення (Elaboration Language-EL).

SL є графічною мовою, спеціально призначеною для представлення компетентними фахівцями в даній області системи основних даних в формі онтологічної інформації (табл. 1.1).

Таблиця 1.1 - Представлення системи основних даних в формі онтологічної інформації

| Елемент онтології                                | Графічне позначення елементу   |
|--|--|
| Позначення класу                                 |     |
| Позначення окремого елементу чи екземпляру класу |     |
| Взаємозв'язок між класами: багато до багатьох    |    |
| Взаємозв'язок агрегування між двох класів        |    |
| Взаємозв'язок узагальнення між двох класів       |    |
| Повільна зміна стану класу чи екземпляра класу   |    |
| Швидка зміна стану класу чи екземпляра класу     |  |
| Миттєва зміна стану класу чи екземпляра класу    |   |
| Позначення процесів                              |  |
| Позначення з'єднань                              |  |
| Позначення перехресть                            |  |

В стандарті IDEF5 передбачені чотири види схем, а саме: діаграма класифікації, композиційна схема, схема взаємозв'язку, діаграма стану об'єкта.

Діаграма класифікації служить засобом логічної систематизації знань, що були накоплені при вивченні системи (рис. 1.1). Існують два види таких діаграм: діаграма строгої класифікації (Description Subsumption - DS) та діаграма

класифікації виду (Natural Kind Classification - NCS). За допомогою DS класифікують логічні об'єкти. В діаграмах NCS інтерпретація властивостей класу є більш загальною.

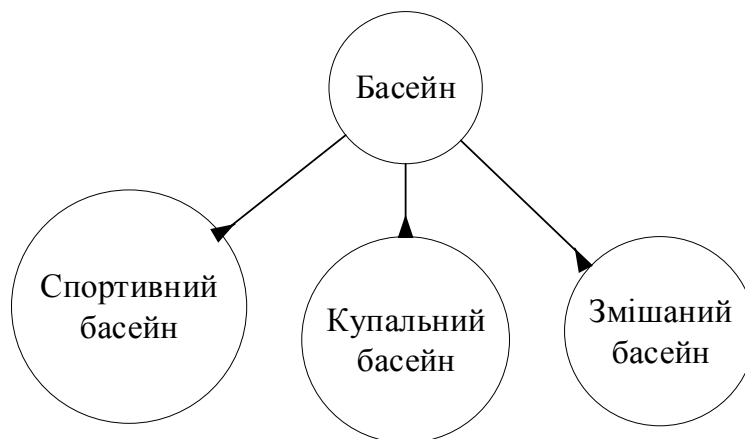


Рисунок 1.1 – Приклад діаграми класифікації

Композиційні схеми (Composition Schematics) використовуються для графічного представлення складу класів онтології. Вони показують відношення «частина–ціле» між окремими частинами онтології (рис.1.2).

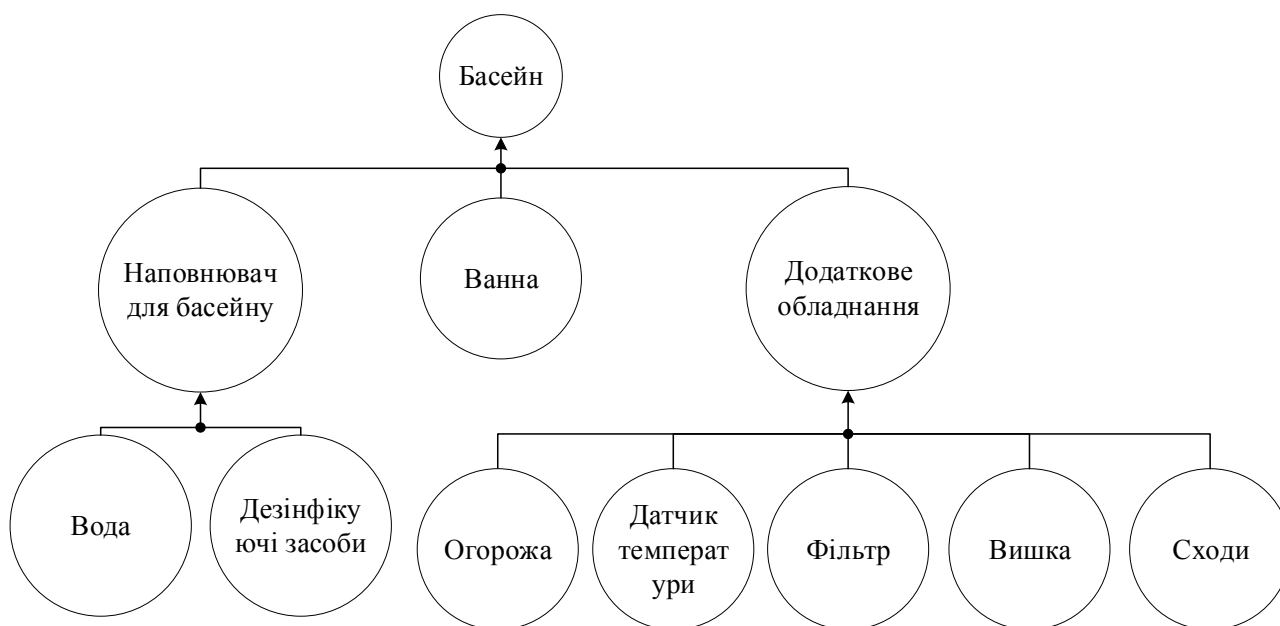


Рисунок 1.2 – Приклад композиційної схеми

Схеми взаємозв'язку (Relation Schematics) дозволяють розробникам візуалізувати і вивчати зв'язки між різними класами об'єктів системи (рис. 1.3-1.4).

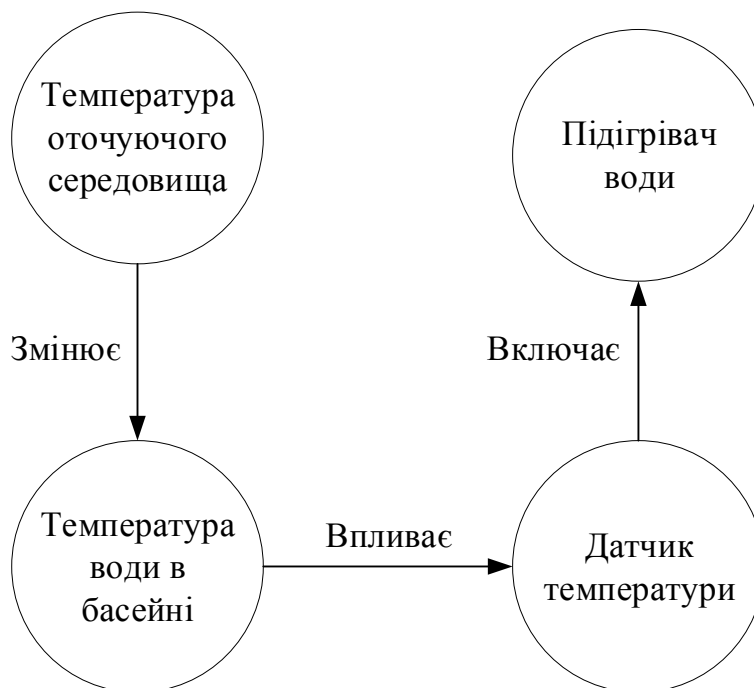


Рисунок 1.3 – Приклад діаграми взаємозв'язку

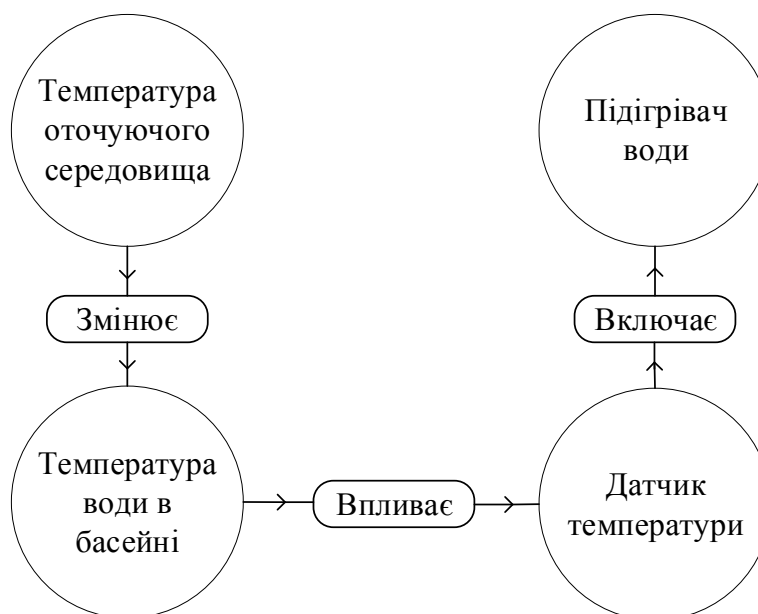


Рисунок 1.4 – Альтернативний вигляд діаграми взаємозв'язку

Діаграми стану об'єкта (Object State Schematics) дозволяють описати процес зміни стану об'єкта. З об'єктом можуть відбутися два типи змін: зміна стану або зміна класу (рис. 1.5).

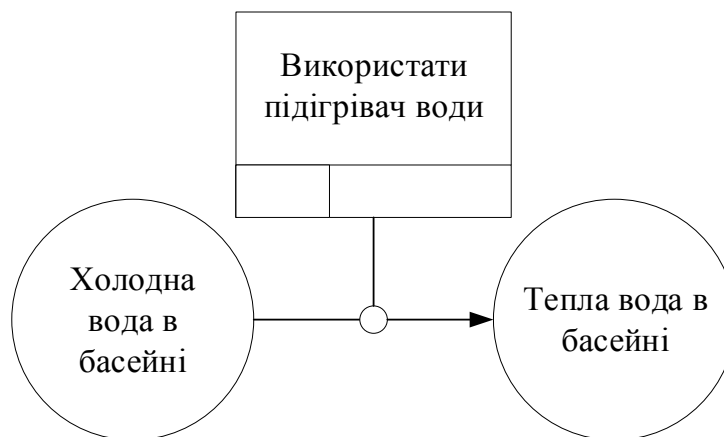


Рисунок 1.5 – Приклад діаграми стану об'єктів

### Хід роботи

1 Задана предметна область – управління відкритим басейном.

Температура навколишнього середовища впливає на температуру води в басейні. Холодні погодні умови призводять до швидкого охолодження води, тому для підтримки належного рівня температури необхідно використовувати нагрівач води в басейні. Велика кількість людей в басейні, які зайняті активними видами водних ігор, також прискорюють охолодження води. Великий натовп ще сприяє тому, що потрібно то додавати воду, то відкривати кришки злива води в басейні для підтримки належного рівня води в басейні.

При аналізі цієї предметної області були виділені наступні характеристики басейну, зміну значень яких можна спостерігати, тобто це перелік вхідних параметрів:

- температура навколишнього середовища [18 °C; 45 °C];
- температура басейну [22 °C; 35 °C];
- рівень води в басейні – 2 м.

Також був розроблений список характеристик басейну з їх значеннями, якими можна керувати для підтримки нормального функціонування басейну, тобто перелік вихідних параметрів:

- включити/виключити підігрівач води;
- відкрити/закрити кришку зливу води;
- включити/виключити подачу води.

2 Згідно заданій предметній області була розроблена фізична модель бази даних (рис. 1.6).

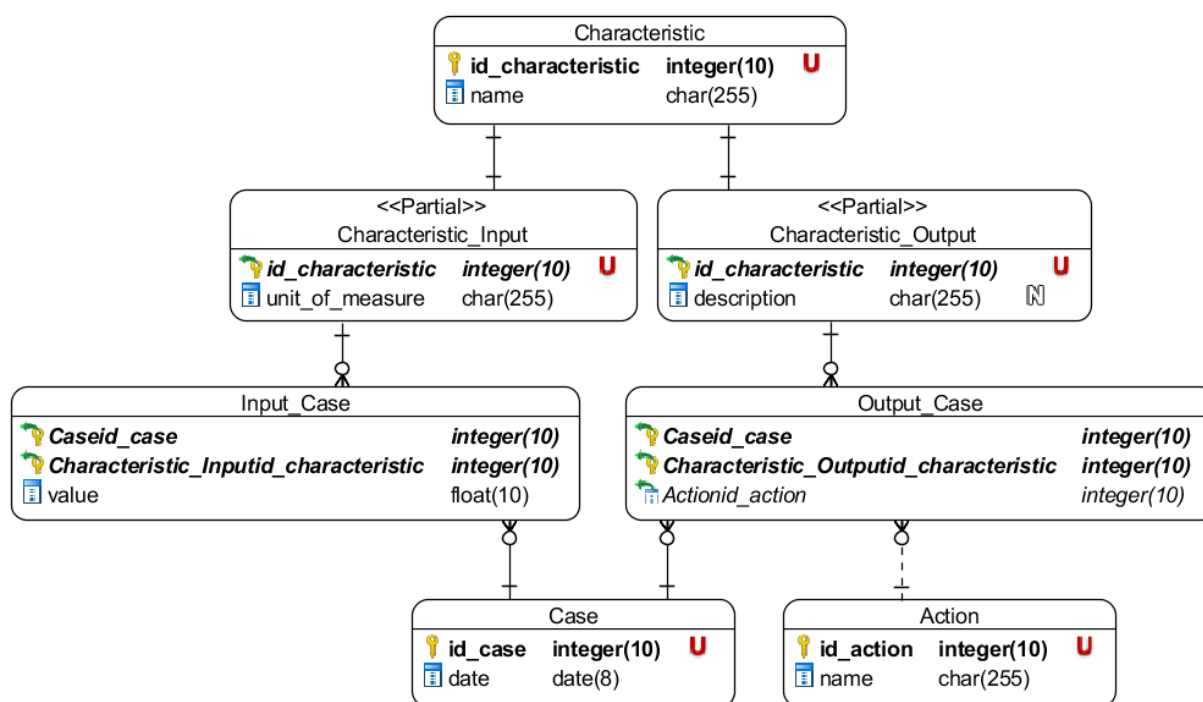


Рисунок 1.6 – Модель бази даних для управління басейном

Перелік сутностей в моделі даних:

- Characteristic – сутність з переліком усіх параметрів басейну, яка розподіляється на дві категорії: Characteristic\_Input – таблиця з набором вхідних показників та Characteristic\_Output – таблиця з вихідними параметрами;
- Action – таблиця, що відповідає за дії вихідних параметрів;



- Case, Input\_Case, Output\_Case – сутності, що зберігають статистику дій при управлінні басейном, тобто при яких значеннях вхідних параметрів як необхідно змінити вихідні параметри.

3 Для обраної предметної області була розроблена онтологія згідно стандарту IDEF5 на прикладі діаграми взаємозв'язку (рис.1.7).

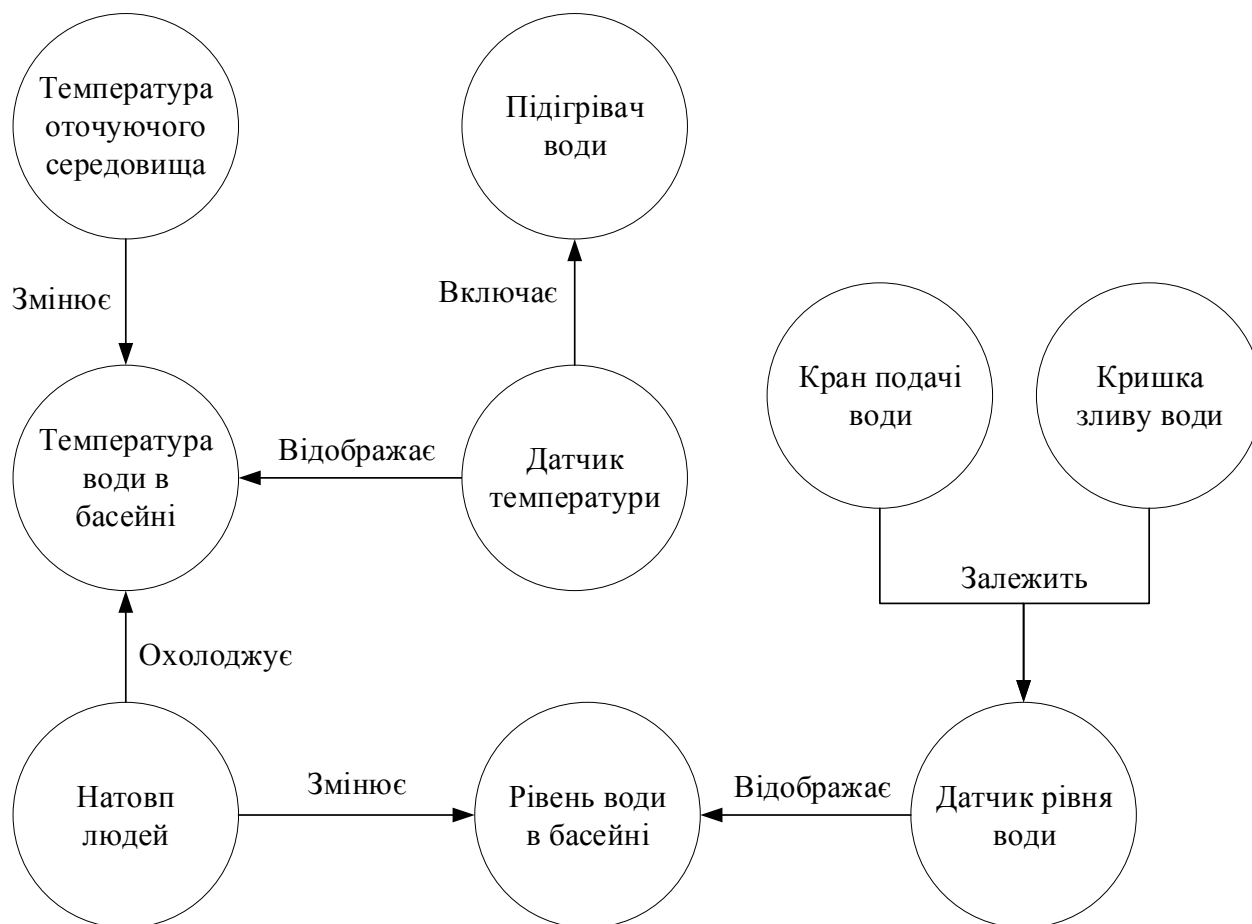


Рисунок 1.7 – Онтологія управління відкритим басейном

Для уточнення процесів між окремими об'єктами необхідно використовувати діаграми стану об'єкта. Наприклад, зменшення температури води приводить до включення підігрівача води, що відображено на рисунку 1.5.

## КОНТРОЛЬНІ ПИТАННЯ

1. Дайте визначення терміну «онтологія».
2. У вигляді яких специфікацій можна графічно представити онтологію?
3. Яке призначення стандарту IDEF5?
4. Які мови використовується при побудові онтологій в IDEF5?
5. Назвіть основні графічні елементи SL?
6. Дайте характеристику основним діаграмам IDEF5.

## ЛІТЕРАТУРА

1. Литвин В. В. Технології менеджменту знань: навч. Посібник / В. В. Литвин. – Львів: Видавництво Львівської політехніки, 2010.-260 с.
2. Верников Г. Стандарт онтологического исследования IDEF5-  
[www.vernikov.ru](http://citforum.ru/cfin/idef/idef5.shtml) <http://citforum.ru/cfin/idef/idef5.shtml>, 11.11.2017
3. Д. И. Муромцев. Онтологический инжиниринг знаний в системе Protégé. – СПб: СПб ГУ ИТМО, 2007. – 62 с.